NOTES 21: CLT-BASED INFERENCE FOR DIFFERENCES

Stat 120 | Fall 2025
Prof Amanda Luby

---

1 CLT Recap

> **Note**
>
> **CLT for 1 Proportion**
>
> $$\hat{p} \sim N(\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_)$$
>
> $$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$
>
> But, we don't know p, so plug in __________ (CI's), __________ (HT's), __________ (conservative)

> **Note**
>
> **CLT for 1 Mean**
>
> $$\bar{X} \sim N(\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_)$$
>
> where $SE_{\bar{X}} =$ __________. But, we don't know σ, so we plug in s. Then,
>
> $$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

Example We want to build a 99% confidence interval for μ, where $\bar{X} = 290$, $n = 120$ and $s = 87.6$. The correct form for the CI is:

2 CLT for a difference in proportions

Idea:

Population parameter:

Sample statistic:

CLT:

$$\hat{p}_1 - \hat{p}_2 \sim N(\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_)$$

How big is "big enough?"

- $n_1 p_1 >$ __________
- $n_2 p_2 >$ __________

What is the SE?

$$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

but, we don't know p_1 or p_2. Three options:

1. "Plug-in" $\hat{p}_1$ and $\hat{p}_2$
2. Use $\hat{p}_{pooled} =$ ________________
3. "Conservative" SE: assume $p =$ ________________

> Note
>
> **CLT-based confidence interval for a difference in proportions:**
>
> $$\hat{p}_1 - \hat{p}_2 \pm z^* \times SE_{plug-in}$$

> Note
>
> **CLT-based test for a difference in proportions:**
> $H_0 : p_1 - p_2 = 0$
>
> $$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE_{pooled}}$$

3 CLT for a difference in means

Idea:

Population parameter:

Sample statistic:

CLT:

$$\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, SE_{\bar{x}_1 - \bar{x}_2})$$

How big is "big enough?"

ℹ Note

Case 1:

- $n_1 >$ ________________
- $n_2 >$ ________________

ℹ Note

Case 2:

- $n_1 \le$ _____ or $n_2 \le$ _____
- AND

What is the SE?

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

but, we don't know σ_1 or σ_2, so we plug in s_1 and s_2 instead:

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

the sampling distribution is no longer a normal distribution, but becomes a ____________________

ℹ Note

CLT-based confidence interval for a difference in means:

$$\bar{x}_1 - \bar{x}_2 \pm t^*_{df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

ℹ Note

CLT-based test for a difference in means: $H_0 : \mu_1 - \mu_2 = 0$

$$t_{df} = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $df =$ _____________________

4 CLT for Matched Pairs

Idea:

Population parameter:

Sample statistic:

CLT:

How big is "big enough"?

What is the SE?

> **Note**
>
> **CLT-based confidence interval for a matched pairs difference:**

> **Note**
>
> **CLT-based test for a matched pairs difference:** $H_0 : \mu_d = 0$